Running head: WHAT STUDENTS LEARN IN PROBLEM-BASED LEARNING

What students learn in problem-based learning: A process analysis

Elaine H. J. Yew

Republic Polytechnic, Singapore

Henk G. Schmidt

Erasmus University, The Netherlands

Correspondence concerning this article should be addressed to Elaine H. J. Yew, School of Applied Science, Republic Polytechnic, 9 Woodlands Avenue 9, Singapore 738964, Singapore. Telephone number: (+65) 65103046, Fax: (+65) 6415 1310, E-mail: Elaine_yew@rp.sg

Abstract

This study aimed to provide an account of how learning takes place in problem-based learning (PBL), and to identify the relationships between the learning-oriented activities of students with their learning outcomes. First, the verbal interactions and computer resources studied by nine students for an entire PBL cycle were recorded. The relevant concepts articulated and studied individually while working on the problem-at-hand were identified as units of analysis and counted to demonstrate the growth in concepts acquired over the PBL cycle. We identified two distinct phases in the process – an initial concept articulation, and a later concept repetition phase. To overcome the sample-size limitations of the first study, we analyzed the verbal interactions of, and resources studied, by another 35 students in an entire PBL cycle using structural equation modeling. Results show that students' verbal contributions during the problem analysis phase strongly influenced their verbal contributions during self-directed learning and reporting phases. Verbal contributions and individual study influenced similarly the contributions during the reporting phase. Increased verbalizations of concepts during the reporting phase also led to higher achievement. We found that collaborative learning is significant in the PBL process, and may be more important than individual study in determining students' achievement.

Keywords: learning processes, problem-based learning, small-group collaboration, self-directed study, verbal interaction

Introduction

Problem-based learning (PBL) can be considered to be a constructivist approach to education. It seeks to create an environment where students learn in the context of meaningful problems, actively construct mental models, co-construct ideas with peers and develop self-directed learning skills in the process (Schmidt and Moust 2000, Norman and Schmidt 1992, Hmelo-Silver 2004). PBL was originally developed in medical schools to help students integrate basic science and clinical knowledge, as well as to develop clinical reasoning and lifelong learning skills (Barrows 1986). However it is now of increasing interest to educators of various levels and disciplines (Gallagher, Stepien and Rosenthal 1992, Kolodner, Camp, Crismond, Fasse, Gray, Holbrook, Puntambekar and Ryan 2003) as it provides a structured framework of active and collaborative learning, in line with current understanding of learning as a constructive and co-constructive activity involving social interactions (Glaser and Bassok 1989, Palincsar 1998).

With increasing implementation of PBL in classrooms, there has been considerable research focusing on the effects of this approach at the curricular level (Schmidt and Moust 2000, Dochy, Segers, Van den Bossche and Gijbels 2003). However, what exactly do students learn in PBL? And how do they learn during group discussions and self-directed study? As argued by Dolmans & Schmidt (2006) and Hak & Maguire (2000), the answers to these questions are still lacking. The goal of this paper is to report on a first attempt to provide a comprehensive account of the learning-oriented activities of students – what they do and say – throughout the PBL learning process, as well as to identify relationships between the contents of the learning activities of students with their learning outcomes.

PBL, as its name implies, always starts with a problem. This problem refers to an academically or professionally relevant issue that students are supposed to learn more about. Students do not prepare for the problem beforehand and therefore begin their initial discussions based on their prior knowledge. As a group, the students analyze the problem, generate possible explanations, build on one another's ideas, as well as identify key issues to be studied further. The purpose of this exercise is to construct a shared initial explanatory theory or model explaining the problem-at-hand (Schmidt, Van der Molen, Te Winkel, & Wijnen 2009). After this period of teamwork, they disperse for a period of self-directed study to work on the learning issues identified. When they next meet as a team, they are expected to share and discuss their findings, as well as refine their initial explanations based on what they have learned. A tutor is present during the team discussions to help facilitate the learning processes. Thus, the cycle of PBL can be seen as being made up of three phases: initial problem analysis, followed by self-directed learning, and a subsequent reporting phase (Barrows 1988, Hmelo-Silver 2004). Described this way, PBL can be considered a constructivist approach to instruction, emphasizing collaborative and self-directed learning, and being supported by flexible teacher scaffolding (Schmidt, Loyens, Van Gog, & Paas, 2007).

A number of studies have focused on the effects of the initial problem analysis on learning. Schmidt, De Volder, De Grave, Moust and Patel (1989), and De Grave, Boshuizen and Schmidt (1996) have demonstrated that students who discussed a problem in a small group before studying a relevant text learned more from that text relative to students who did not have the chance to discuss the particular problem. It was found that the opportunity for students to elaborate on what they know or think about a subject before studying relevant resources, helped them remember the concepts learned better. In another study, Capon and Kuhn (2004) compared problem-based discussion with expository teaching. They found that students who experienced PBL demonstrated superior explanations and understanding as compared to students in the lecture/discussion group. Their results support the hypothesis that the advantages of PBL over traditional lecture-based instruction lies in its ability to help students integrate new concepts with existing knowledge. While this study suggests a possible mechanism of how PBL enhances understanding, it does not provide us with information on how the students' verbal interactions during the 3 hour session studied helped students learn in the process.

The study by De Grave et al. (1996) explored the relationship between students' verbal interaction and their cognitive change during the problem analysis phase of PBL. Their findings indicate that although verbal interaction represents only a portion of the cognitive processes taking place in a student, it *does* reflect the theory-building processes in the students' learning. This suggests that observational studies of the PBL tutorial process can provide valid data regarding students' thought processes involved in learning. It also suggests that the quality of what students articulate is likely to be related to their learning outcomes. However, De Grave et al. do not report any learning outcomes.

In a study on students' interaction processes, Visschers-Pleijers, Dolmans, de Leng, Wolfhagen and Van der Vleuten (2006) investigated time spent on different types of verbal interactions during group discussion sessions. Using video-recordings of four tutorial group sessions, they found that time spent on learning-oriented verbal interaction was very high- around 80%, and also identified how different types of verbal interactions such as cumulative reasoning, exploratory questioning, and handling of "cognitive conflicts" were distributed over the group meetings. However, no relationships with the amount and content of subsequent learning were reported.

A naturalistic study of the PBL process by Koschmann, Glenn and Conlee (1997) involved a description and analysis of a segment of a PBL meeting up to the point where a learning issue was generated in order to demonstrate the actual events taking place in a PBL tutorial. Another publication by the group focused on a short segment of a PBL group's interaction to provide insight into how students interact in the process of presenting one's own theory and responding to those of others (Glenn, Koschmann and Conlee 1999). One limitation of these studies is that they only examined specific portions of the PBL tutorial and did not relate the behavior of students to later achievement.

The studies cited above all focus on the problem analysis and reporting phases of PBL. Research dealing with the phase of individual, self-directed study is scarce. One study involving the self-directed learning phase was carried out by Dolmans, Schmidt and Gijselaers (1995). They investigated the relationship between student-generated learning issues during problem analysis phase of a PBL classroom and what students actually studied during self-study time. Even though it is generally assumed that students would make use of the learning issues generated to determine their learning activities during self-directed study, this turned out to be the case only to some extent. It seemed that the learning activities of the students during the self-study phase were also determined by other factors such as the nature of tutor guidance and the learning resources available. However, since the measurement of what students actually were studying was based on retrospective self-report, the results may have been biased to some extent. Another study by Van den Hurk, Wolfhagen, Dolmans and Van der Vleuten (1999) focused on how students made use of their self-study phase in terms of learning issues previously generated and time spent on individual study. They found that higher-year students were more self-directed learners compared to first year students, and that those who studied beyond the learning issues generated by the tutorial group during problem analysis phase showed better achievement in tests. This study also relied on self-report data.

Although a fairly detailed picture regarding the nature of the discussions in the tutorial teams has emerged from these studies, there is a definite lack of research investigating what students actually do during self-directed study and how their activities influence the outcomes of their learning. Although in particular the experimental studies discussed above (Schmidt, De Volder, De Grave, Moust and Patel 1989, De Grave, Boshuizen and Schmidt 1996, Capon and Kuhn 2004) suggest that group discussion and elaboration play an important role in students' learning in PBL, the way in which they affect learning remains to be clarified. Moreover, although studies as described above have provided some insight on how group processes and self-directed study (separately) affect student achievement, research on the function and influence of *both* collaborative learning and self-directed study in the PBL process is still lacking. This seems strange given that both phases are believed to be essential aspects of PBL. Without understanding their mutual influence and how they each affect student learning, important questions about how PBL works remain unanswered. As noted by Capon and Kuhn (2004), it is critical to examine the various components of PBL to ascertain what is and what is not essential for PBL to take place. For example, can students learn just as effectively if they were to focus on self-directed learning and research, without collaborative problem analysis and discussion, and vice versa? How, if at all, does each phase of PBL influence the next? Which influences students' learning to a greater extent – collaborative group discussions or individual self-study? As argued by Hak and Maguire (2000), lack of understanding regarding the relationships between these learning processes of PBL with students' learning outcomes means that there is "no research base for giving pertinent advice to students and tutors about how to conduct PBL tutorials (p. 771)". Furthermore, research required to uncover which aspects of the tutorial process are crucial to students' learning should be focused on the actual learning activities occurring in the various phases of PBL (Dolmans and Schmidt 2006, Hak and Maguire 2000). Yet such direct-observation studies are rare as most studies on the activities that occur in PBL were conducted using student self-report, which may be biased to some extent (Schmidt and Moust 2000, Van den Hurk, Dolmans, Wolfhagen and Van der Vleuten 2001).

One observational study that did focus on the learning-oriented interactions of students was conducted by Yew and Schmidt (2008). Here the verbal interactions taking place during an entire PBL process were analyzed qualitatively and the results demonstrated that PBL stimulates constructive, self-directed, and collaborative learning processes as defined in the relevant literature. However, no relationships between the content of their interactions with subsequent learning were reported.

The first purpose of the present study, therefore, was to increase our understanding of the learning processes in all the phases of the PBL cycle, including the self-directed study periods. To this end, we examined whether it was possible to use students' verbalizations and computer resources to provide an account of their ongoing learning. Secondly, we aimed to identify the relationships between the actual learning activities of students and their learning outcomes using data from "on-line" observations. The studies reported here are, as far as we know, unique in that they have taken place in an educational context where the self-directed learning phase can be observed in its natural classroom setting, allowing for insights into how students learn during this time. In the polytechnic where the studies were carried out, the problem analysis, self-directed learning and reporting phases of PBL all occur within one day. All students have a personal laptop that can be connected to the internet. First-year students generally rely mainly on internet resources for their research, and remain in class for their self-directed study. Thus it was possible to record and hence observe students' verbal interactions and internet activities for the entire PBL process, even during the periods of self-directed study.

Furthermore, unlike previous studies (e.g. Capon and Kuhn 2004), the present study focused on the *contents* of what was discussed and learned, and attempted to relate these contents to subsequent achievement. To that end, we recorded group discussions about a problem in genetics, logged all individual study activities of the students while they were using online resources, and recorded contributions of the facilitator. In addition, we measured prior knowledge of the subjects and actual learning gains. The resulting protocols were analyzed to understand the growth in usage and study of relevant concepts over the different learning phases and its effect on achievement. The units of analysis were the scientific genetics-related concepts or terminologies that the students articulated and studied while gaining insight in the problem at hand. Concepts such as “DNA,” “alleles,” or “meiosis” can be considered micro-theories (Murphy and Medin 1985) that students use in the course of trying to learn about genetics. We hypothesized that the frequency with which these concepts were used by students while discussing the problem and studying subject-matter, could be considered an indicator of the learning-oriented activities going on and would determine subsequent achievement. While we do not mean to say that the usage of a concept in itself implies understanding, we hypothesize that increased usage over time would be the result of increased learning. We believe that, in the process of learning, the students gradually build up a cognitive representation of the explanations for the problem at hand consisting of a semantic network of concepts related by links. The more students have learned about a topic, the richer and more detailed this particular network would be (Glaser and Bassok 1989). The richer the semantic network, the more concepts the student has available to understand the issues at stake. Hence measuring the number of relevant concepts articulated and studied by the students in each learning phase gives an indication of the adequacy of these students' learning. In this paper, we report two separate studies. The first study involved nine participants collaborating in two groups. We followed these two groups throughout the day, resulting in roughly 16 hours of discussion protocols and 70 hours of individual internet log files. The need to conduct a fine-grained study of learning in PBL precluded involving more students. However as the small sample size limited the types of analyses that could be performed, a second study involving 35 students was later carried out. This resulted in verbal interactions that amounted to about 60 hours and computer screen recordings of approximately 260 hours. The resulting protocols were analyzed to identify the number of concepts relevant to the learning issues related to the problems.

In summary, the research questions addressed were: (1) how are the numbers of relevant concepts verbalized and studied from resources (and the frequencies of their occurrence) distributed over the different learning phases of the PBL process? What does this distribution of concepts over time suggest about the nature of the learning process in a problem-based curriculum? (2) How is student achievement influenced by their verbalization and by their individual study in the different phases of PBL? How do collaborative learning and self-directed study in the PBL process influence each other, if at all? Which (if either) plays a bigger role in predicting student achievement? And finally, (3) can students' learning-oriented verbalizations and their computer resources log help provide an account for ongoing learning in the PBL process?

Study 1

Method

Participants

Participants were nine first-year students from a polytechnic in Singapore. In this polytechnic, all first-year students undergo a common curriculum regardless of their subject discipline. The nine students (making up two different teams) were from the same Basic Science class and were being recorded on the fifth week of Semester Two. Students were not new to PBL as they had already completed Semester One (16 weeks of PBL classes from Monday to Friday). The facilitator had several years of experience. Both students and facilitator gave informed consent.

Educational context

The PBL process in this polytechnic is unique in that a "One-day-one-problem" approach where students work on one problem in a day is adopted. It takes place in a class setting consisting of 25 students and one facilitator. The students are grouped into teams of approximately five students. The daily routine consists of three meetings with facilitator interaction and two periods of self-directed study or teamwork without facilitator involvement. A brief description of the day's process is shown below:

- Phase 1: Problem analysis phase (approximately 1 hour): Facilitator presents problem for the day. Students work in teams of five to identify their prior knowledge and learning issues.
- Phase 2: First self-directed learning (SDL-1) period (approximately 1 hour): Students do individual research or work with their teams on worksheets and other resources provided. Time is spent teaching one another within the team. Most of the individual research is done by reading online resources from the internet.
- Phase 3: Group meeting with facilitator (approximately 1.5 hours): Each team of students meets with the facilitator for about 20 minutes to share their progress and strategy of understanding the problem. The rest of the time is spent continuing on self-study and/or discussion.
- Phase 4: SDL-2 (approximately 2 hours): Extended time where teams consolidate their research and formulate a response to the problem.
- Phase 5: Reporting phase (approximately 2 hours): Each team presents their consolidated findings and response to the problem, defending and elaborating based on questions raised by peers and the facilitator. The team presentation is usually in the form of powerpoint slides. The facilitator would also clarify key ideas if necessary.

Although the PBL approach in this context has been adapted to suit the learning needs of the particular students and takes place within one day, it is to be classified as PBL based on the 'six core characteristics of PBL' as described by Barrows (1996). These characteristics include student-centered learning with small groups working under the guidance of a tutor who acts as a facilitator. Students work on authentic problems with no prior preparation so as to achieve the required knowledge. In addition, no direct teaching takes place; all learning is student generated. Finally, it is through self-directed learning that students acquire new information (Hmelo-Silver 2004). Dochy, Segers, Van den Bossche, and Gijbels (2003) used the same criteria in their meta-analysis on the effects of PBL for inclusion in their study.

Procedure

The two teams involved in the study had been working together in the Basic Science class for two weeks before being recorded. Although only two teams were being observed, there were a total of four teams in the facilitation room for that day.

Verbal interaction was recorded using a digital audio recorder placed at each team's table. Students' computer usage was tracked using the program Camtasia Studio Screen Recorder (TechSmith Corporation, Okemos, MI) installed on each student's laptop. The students were audio-recorded and had their computer usage recorded twice beforehand to allow them to be familiar with the procedure on the actual day. Both recording devices (voice and computer) were started at the beginning of the day when the problem was first shown to the students till the end of the day when the facilitator concluded the class (a total of about 8 hours). To ensure an authentic recording that was representative of what usually happens during the self-directed study times, no facilitator or observer was present during the SDL sessions. The room in which the recordings were carried out was the students' regular classroom.

Materials

The problem statement for the day was entitled "Code of Life" and it introduced students to the concept of heredity and genes. Students were to explore the role and properties of the gene molecule which is able to transmit information from parents to children. The problem is part of a Basic Science module which aims to introduce foundational scientific principles and applications for all students in the polytechnic, regardless of their specific discipline of study. This is a general module that includes a wide range of concepts such as energy, electricity, atomic structure, structure of organic compounds, cells, recombinant DNA technology, Newtonian mechanics and special relativity. The problem is presented in Appendix 1.

A concept recognition test was administered immediately after the day's problem to gauge students' learning achievement. This test is a simplification of the concept mapping technique (Novak 1998) and consisted of a list of 34 concepts that are more or less closely related to the central topic of heredity. This set of 34 concepts cover the domain as a whole and was based on an exhaustive review of relevant literature from textbooks and internet resources. Examples of such concepts include 'gametes' and 'chromosomes'. A number of concepts not related to heredity ("fillers") were interspersed in the list. Examples include 'water' and 'oxygen'. Students were instructed to rate the extent to which the concepts listed were related to heredity using a scale of 1 to 5, where 1 = not at all related; 2 = a little bit related; 3 = to some extent related; 4 = quite closely related and 5 = very closely related. No time limit was set for the test.

Two colleagues with expertise in the field of molecular biology were asked to identify the most appropriate answers independently. Inter-rater agreement was 83.8%. Where there were differences in rating, a third opinion from a similar expert was sought. Student rating of each concept scored 2 points if it was the same as the expert answer, 1 point if it differed by ±1, and 0 for any other answer.

Analysis

Verbatim transcripts of a total of 16 hours of verbal interactions of the two teams were produced. The computer screen recordings of the students were viewed to identify the websites they had accessed. This amounted to around 70 hours of screen recording as each student was online for about 7-8 hours when working on the day's problem. The unit of analysis was the number of relevant propositions or concepts related to the theme of heredity, as identified by an exhaustive review of literature. The concepts articulated and studied from online resources were counted for each student for each learning phase (i.e. problem analysis, SDL-1 etc). The *total frequency* of concepts refers to the total number of relevant concepts verbalized or studied, including those that were repeated in one phase. On the other hand, the total number of *different concepts* did not include those that were repeated during the same phase. *Newly emerged concepts* were those that were not previously mentioned by the individual in any prior learning phase of the day.

The following excerpt from a discourse of one team during the problem analysis phase (Phase 1) is shown to demonstrate how the relevant propositions verbalized by each student were counted.

Excerpt taken from initial part of problem analysis phase (Phase 1):

> (The facilitator has just given the teams 10 minutes to come up with relevant keywords or ideas related to the problem statement.)
>
> L: Ok, what's a gene?
>
> C: DNA or something like that…
>
> L: Ok, got some key words right- like chromosomes, X and Y chromosomes.
>
> S: Chromosomes, alleles, X and Y chromosome, phenotypes, genotype, dominant, recessive…
>
> L: I only learned allele.
>
> SF: What about gametes?
>
> C: Cheem… (Nb: 'Cheem' is a Hokkien term which means "complex" or "difficult to understand")
>
> L: What are gametes?
>
> S: Gametes are your sperm and your egg.

The relevant concepts counted for this excerpt would include 'gene', 'DNA', 'chromosomes', 'X and Y', 'alleles', 'phenotype', 'genotype', 'dominant', 'recessive', 'gametes' and so on. The total number of relevant concepts articulated by student L would include one count of ‘gene’, two counts of ‘chromosome’ and one count of ‘alleles’ and so on. However the number of different relevant concepts uttered would only count ‘gene’, ‘chromosome’ and ‘alleles’ once each. If L were to mention the proposition ‘gene’ again in say, SDL-1 (Phase 2), it would still be counted once as the number of different relevant concepts uttered for that meeting. However for the counting of newly emerged concepts, any concepts mentioned here would not be counted again in subsequent phases.

The amount of time spent on-task was deduced from the students’ computer screen recordings as well as the audio recordings.

One-way ANOVA was used to find out if there were significant differences in the mean number of concepts verbalized or studied during each learning phase. Correlation analysis was performed to examine the relationships between students’ learning outcomes with the following: verbal interactions, self-directed learning, time spent on-task and facilitator’s contributions.

Results and Discussion

As mentioned earlier, the concepts verbalized by students during the different learning phases were counted in three different ways. The distribution of the average number of *total* relevant concepts verbalized for each of the learning phases of the PBL process for the nine students is shown in Figure 1. Second, repetition of concepts within a learning phase was excluded in the count. Figure 2 is a distribution of the average number of different concepts verbalized for each learning phase of the day by each team. Lastly, only completely new concepts were counted. The average number of these newly emerged concepts for each learning phase is shown in Figure 3.

(Insert Figures 1, 2 and 3 here)

The one-way ANOVA revealed that the concepts verbalized differed significantly as a function of the different learning phases: for total number of concepts, $F(4, 40) = 3.40$, $p < .05$; for number of different concepts within each learning phase, $F(4, 40) = 3.68$, $p < .05$ and for number of newly emerged concepts for the day, $F(4, 40) = 3.64$, $p < .05$.

Post-hoc analyses using the Games-Howell test showed that the number of newly emerged concepts verbalized was significantly higher for the group meeting with facilitator (Phase 3) ($M = 8.3$, $SD = 4.39$) as compared to SDL-2 (Phase 4) ($M = 3.0$, $SD = 2.06$) ($p < .05$) while the number of different concepts articulated during Phase 3 ($M = 17.0$, $SD = 7.43$) and reporting phase (Phase 5) ($M = 16.0$, $SD = 6.58$) were significantly higher than for SDL-1 (Phase 2) ($M = 6.9$, $SD = 3.48$) ($p < .05$).

Relevant concepts accessed via online resources were counted similarly. The distribution of the average number of total relevant concepts, different concepts within each learning phase and newly emerged concepts for the day that were accessed online for each of the learning phases are shown in Figures 4, 5 and 6 respectively.

(Insert Figures 4, 5 and 6 here)

The one-way ANOVA revealed significant differences in the numbers of concepts accessed online for the five learning phases: for total number of concepts, $F(4, 40) = 2.75$, $p < .05$; for number of different concepts within each learning phase, $F(4, 40) = 4.42$, $p < .05$ and for number of newly emerged concept for the day, $F(4, 40) = 4.85$, $p < .05$.

Post-hoc comparisons using the Games-Howell criterion for significance showed that the number of newly emerged concepts accessed online was significantly higher for problem analysis phase (Phase 1) ($M = 12.7$, $SD = 10.48$) and group meeting with facilitator (Phase 3)

(*M* = 13.4, *SD* = 10.54) compared to reporting phase (Phase 5) (*M* = 0.56, *SD* = 1.13) $(p < .05)$ while the number of different concepts accessed online during group meeting with facilitator (Phase 3) (*M* = 25.3, *SD* = 10.25) was significantly higher than for Phase 5 (*M* = 5.1, *SD* = 7.87) $(p < .05)$. No other specific post-hoc contrasts were significant.

The results above show that the highest number of newly emerged concepts occurred during the problem analysis phase (Phase 1). This suggests that many of the relevant concepts are already known in one way or another and that the initial discussion of the problem serves to reactivate those concepts. However, the highest total number of relevant concepts and number of different concepts articulated by the students occurred during the group meeting with facilitator (Phase 3). This suggests that this learning phase is particularly rich in terms of articulation and repetition of the concepts learned previously. Interestingly, the total number of concepts studied online was *also* the highest during Phase 3, whether it was the total number or the number of different concepts or the number of newly emerged concepts that was being considered. Thus, it appears that students' research activities as well as verbal interactions were highest during this period. This phase is when the facilitator meets with each team to find out their progress. Since the facilitator only spends about 20 minutes with each team during this time, results here indicate that students spend the remaining time of about 45 minutes to 1 hour on-task, researching individually and discussing as a team. Thus it can be seen that in the present context, the distinctions between student-facilitator interaction, group discussion and self-study phases are blurred. This is due in part to students having their personal laptop computers with them throughout the day and are able to continually access internet resources, whether it is the discussion or SDL phase of the PBL process.

Excerpts of the two teams taken from Phase 3 are shown below to demonstrate the contexts in which students articulate relevant terminologies. In Team 1, students A and SF are co-constructing their understanding about RNA. It can also be seen from the example of

Team 2 that the students tend to share what they have read up from online resources and in the process of further discussion, make sense of the new ideas they have read.

Team 1

A: So RNA is actually information within the DNA

SF: Mm, actually DNA transcribes the RNA, then RNA produces the peptide chain to become protein

A: So it's the cell reads DNA and then-

SF: Actually RNA is part of DNA… (unclear)

A: So it's like the messages… like the message within the DNA

SF: Yeah

A: And then RNA and then the cell produces protein

Team 2

ZW: We have to find out more about genes, right? About how does it play in the organism

J: It says that genes produce all the protein

ZW: Yeah

J: Control all your proteins

ZW: Yeah yeah yeah correct. Yeah I read that. You go to worksheet, first question, there is a link already.

J: No I read from this. Easier. 'cos animation.

ZW: Blood contains lots of red blood cells that transport oxygen for our body. The cells use the proteins called haemoglobin to capture and carry the oxygen. From our 40000 genes, only a few contain instructions for making haemoglobin proteins. The remaining genes contain instructions for making other parts of our body

J: That means we have 50000 genes ah?

Z: No forty over thousand

…

ZW: Wow. Cool. J: You go to this website. It's easier to understand I think.

The data shown in Figures 1 to 6 also give insight into what students do during the self-directed study periods. It can be seen that generally the new concepts verbalized emerge in the problem analysis (Phase 1) and SDL-1 phase (Phase 2), while there is a greater number of repetitions in the group meeting with facilitator (Phase 3) and SDL-2 period (Phase 4). A similar trend occurs for the concepts accessed online. This suggests that in the first phases of the learning process in PBL, students focus on “initial concept articulation” while in the later phases of the process they focus on “concept repetition”. Although we did not do a qualitative content analysis of the students’ discourse, a brief excerpt from a typical episode of verbal interaction during this learning phase demonstrates to some extent the elaboration and repetitions involved in the process.

> L: Can just give me a brief explanation of meiosis?
>
> SF: Meiosis-
>
> A: Meiosis simply-
>
> SF: Is a-
>
> A: I try to explain- meiosis is simply the division of chromosomes but is different from mitosis as in, you know we have 46 chromosomes … so in mitosis rite they actually duplicate themselves
>
> L: Mm hm
>
> A: And then split into two and then both have 46 chromosome
>
> L: Mm hm.
>
> A: But meiosis it actually divides itself and has 23 chromosomes only- so 23 chromosomes are just for reproduction, so-
>
> L: Oh for reproduction of the same cell…
>
> …
>
> SF: You say after meiosis rite, it become mitosis
>
> A: Yeah, because just now I was trying to say that meiosis-
>
> SF: Yeah
>
> A: The main reason it happens is because- for the reproduction of the baby. So when-
>
> SF: When the- so like they join together
>
> A: Then mitosis starts
>
> SF: The er… the zygote will undergo mitosis

A: Mitosis

SF: To split into many cells

A: Yeah

Furthermore, a comparison of the trends in Figures 1 and 4 suggests that what students do during the two SDL periods (Phases 2 and 4) differ to some extent: the first phase involves less verbalization of concepts and more research, while students engage actively in both discussion and research during the second period.

Finally, Figures 4 to 6 show that the number of relevant concepts accessed during the reporting phase (Phase 5) to be very low. This is to be expected as students would be listening to other teams present their findings during this reporting phase, and unlikely to still be actively searching for relevant resources to study.

Thus the first conclusion of this study is that in the process of PBL studied here, two different phases can be observed – an initial *concept articulation* phase - consisting mainly of the problem analysis phase and SDL-1, and characterized by the emergence of new concepts articulated and studied online, and secondly, a *concept repetition phase* (mainly the group meeting with facilitator phase and SDL-2) where relevant concepts are repeated. We also found that the most extensive on-task activity occurs half-way in the process, during the meeting with facilitator (Phase 3); in this phase most verbal interaction and online research were taking place. It seems that students first need a certain period “to warm up” before they deeply engage in the study of subject-matter. This could also indicate that most of the students do not fully utilize the first period of self-directed study, tending to spend it on off-task matters and getting more serious about the task-at-hand when the facilitator is present in class during Phase 3.

We will now present and discuss the results for the second aim of this study investigating how student achievement is influenced by their verbalization and by their individual study. Since the number of students studied was limited, multiple regression analysis of verbalizations in the different phases of the learning process on the learning outcomes was not considered meaningful. Therefore simple descriptive correlation analysis was conducted. The results of correlation analysis between students' learning achievement scores and the total number of relevant concepts articulated during the different learning phases are shown in Table 1. The results with respect to concepts uttered during the verbal interactions were as follows: The total number of relevant concepts articulated during the reporting phase was significantly correlated with students' concept recognition test scores ($r = .83$, $p < .01$). Similar correlations were found when the number of *different* concepts articulated was considered instead. In addition, the total number of different concepts for the day (including all phases) correlated with students' achievement ($r = .59$, $p < .01$) as well. This shows that students who use *more* different concepts in the discussions over time also learn more. Thus our findings show that total repetition of concepts in particular during the reporting phase, as well as the breadth of terminologies articulated throughout the day play significant roles in the learning process. No significant correlations were found between the number of newly emerged relevant concepts and student achievement.

The observation that both the total number and the number of different relevant concepts articulated during the reporting phase (Phase 5) correlate strongly with students' learning indicates that students who have learned tend to share more of their findings and understandings during the reporting phase. This is a useful finding when assessing student progress at the end of the PBL process as it indicates that the amount of information articulated by the students at the end of the PBL process gives a good representation of their learning gains.

The results of correlation analysis between students' test scores and the total number of relevant concepts encountered online during the different learning phases are shown in

Table 2. Significant correlations were found between the total concepts studied online during the SDL phases and student achievement suggesting that individual research on online resources during SDL was important to student learning. The amount of study during the first SDL phase appeared to play a greater role in predicting students' final learning outcomes compared to the second SDL phase as the correlations between student learning and test scores were statistically significant for the former while not for the latter. The total number of concepts accessed online during the problem analysis phase was also significantly correlated with student achievement.

Similar statistically significant correlations were found for the number of *different* concepts studied online during SDL times as a whole with students' test results, with the breadth of concepts studied in the first SDL phase being of particular importance. These results show that both the amount and breadth of self-directed research undertaken from the study of online resources are highly indicative of students' learning from the PBL process.

In the case of newly emerged concepts accessed from online resources, there was a significant positive correlation between total numbers of new concepts read throughout the day with the concept recognition test scores. Significant positive correlations also exist between the number of new concepts read in the studied in the first SDL phase and test scores. Interestingly, a significant negative correlation exists between newly emerged concepts read during the group meeting with facilitator (Phase 3) with the concept recognition test scores. These results indicate that the more new concepts accessed or read during the earlier part of the learning process was related to student achievement, while coming across new concepts only towards the later part of the day was indicative of lesser learning by the end of the day. Students who were less on-task during the first period of self-directed study would naturally have more newly emerged concepts when they read online resources during Phase 3 compared to others who had made use of the self-directed study time to acquire new concepts.

Thus, these results show that students who made use of the initial knowledge acquisition phase to research and uncover new concepts in the process achieved more learning at the end of the day. This deduction is in line with the finding that time spent on-task correlated significantly with results for the concept recognition test (r = .64, p< .05). Since all the students were generally on-task during meeting times when the facilitator was present, time spent on-task was relevant mainly during the self-study periods.

The other element of the PBL-process which also contributed to learning was the extent of the facilitator's contributions (r = .62, p < .05). This result is particularly striking because the facilitator's interventions and contributions were limited in number. As our current data cannot be used to determine the cause-and-effect relationship between facilitator's verbal interaction and students' learning, further investigation is necessary to better understand the effect of facilitator contribution to students' learning in the PBL context.

Thus in regards to the question of how student achievement is influenced by their verbalization, we report that between the repetition and the range of different concepts articulated, it is the latter that plays a greater role in student learning.

In the case of students' self-directed study, it was revealed that being exposed to a greater number of different relevant concepts as well as having increased exposure to the same concepts correlated with students' learning outcomes. Comparing this result with that of the correlations of the total number of students' verbal interactions during SDL with achievement (r = .20, p > .05) suggests that individual study plays a greater role in their learning than verbal interactions with peers during the SDL phases.

Conclusions of Study 1

In conclusion, in regards to the first objective of our study, we have found that the process of PBL is characterized by two distinct phases: an initial phase in which students focus on the acquisition (and use) of new concepts and a subsequent phase in which they mainly repeat, or elaborate upon, these previously acquired concepts. The implications of this finding will be discussed further in the general discussion.

We have also shown that what students do during self-directed study plays the key role in their learning outcomes. Studying a greater number of relevant concepts as well as having increased exposure to the concepts had strong correlation with students' achievement. The total number of concepts students articulate correlate slightly less strongly with their learning, although the correlation becomes more significant when the number of different relevant propositions (i.e. breadth and variety of relevant concepts) is considered instead of the frequency (repetition of relevant concepts). However due to the limitations as a result of our fairly small sample size ($N = 9$), we were not able to identify causal relationships between what students say and do, with their learning gains. On the other hand, we did find statistically significant effects on learning despite the small size of our sample. Given the fact that the power of the statistical tests used was extremely small due to sample size, the significant results are the more telling. They suggest that our findings are meaningful and likely to be valid. We therefore decided to replicate the study with a larger sample size so as to carry out more detailed analyses using structural equation modeling. This would enable us to uncover causal relationships between students' verbal interactions, individual study, and student learning, as well as to study the influences exerted by results of one learning phase on the next.

Our findings in this study indicate that both verbal interactions and self-directed study correlated with students' learning outcomes. Social constructivism also suggests that knowledge is constructed by means of collaborative interactions (e.g. Cobb 1994, Driver, Asoko, Leach, Mortimer and Scott 1994), while research on self-regulated learning has shown that the use of self-regulated learning strategies strongly influences academic achievement (Zimmerman 1990). We therefore hypothesized that both small-group collaboration and individual study influence achievement. Since learning is generally thought to be a cumulative and constructive process where new learning builds upon knowledge acquired in a previous phase, we hypothesized that the relevant concepts students articulated during the problem analysis phase influences that in the self-directed study phase and finally the reporting phase. Finally the concepts articulated during the reporting phase would influence their learning achievement directly. As studies have indicated that small group discussion stimulates students' interest in the subject matter (reviewed by Dolmans and Schmidt 2006) we also expected that the extent of articulation of concepts during problem analysis would influence the amount of individual self-study occurring during the self-directed study phase which would further influence the articulations of concepts during this phase and the reporting phase. Figure 7 summarizes our hypothesized relations in terms of a causal model.

(Insert Figure 7 here)

The following section describes Study 2, involving a similar methodology but using a larger sample size in order to test the hypothesized causal model using structural equation modeling.

Study 2

Method

Participants

Participants were 35 first-year students from the same polytechnic in Singapore. The 35 students were from two different Basic Science classes and were being recorded on the seventh week of Semester One. Students were not new to PBL as they had already completed

6 weeks of PBL classes (from Monday to Friday) prior to this. Students and facilitators gave informed consent.

Educational context

The education context is the same as that described in Study 1. For Study 2 however, we combined the contents of Phase 2, 3 and 4 as described in Study 1 into one self-directed learning period. This was because only 20 minutes out of the 1.5 hours of Phase 3 was spent with the facilitator, while the rest was self-directed study time, similar to that in Phases 2 and 4.

Procedure

The same procedure as described in Study 1 was carried out.

Materials

Similar to the problem statement in Study 1, the problem statement for the day was also taken from a Basic Science module. Entitled “Heart Matters”, it introduced students to the concepts of the circulatory system and blood pressure. To measure students’ achievement, a concept recognition test as described in Study 1 was administered at the end of the day after the learning had taken place.

Analysis

Verbatim transcripts were analyzed similarly as previously described. This time a total of approximately 56 hours of verbal interactions and 245 hours of computer screen recording for the eight teams were produced. The unit of analysis was again the numbers of relevant concepts articulated during discussion or encountered while engaged in self-study. Their relevance was ascertained by comparing the student-produced or encountered concepts with an exhaustive list of concepts as identified by a review of the literature on the theme of blood transport around the body. In Study 2, we only considered the *total number* of concepts verbalized or studied, including those that were repeated as Study 1 had demonstrated that the other distinctions had only limited added value.

Verbal contribution during problem analysis phase was measured using the total number of relevant concepts uttered by each individual student during the problem analysis phase, while verbal contribution during self-directed learning phases was measured using the total number of relevant concepts uttered during those phases. As described earlier in the ‘education context’ section, student activity during the self-directed learning phases is not limited to individual study and search, but also includes group discussion, peer consultation and collaboration. The extent of individual search and study conducted by students was estimated using the total number of relevant concepts accessed online during the self-directed learning phase. We are aware that this method would not distinguish between what students only browse through and what is actually read and studied. However, we consider it a “best guess” about what is learned from websites. Alternative methods, such as asking students as to what is learned, would intrude upon the students’ learning process and perhaps distort it; our recording method does not. Furthermore, results from Study 1 shows that the number of concepts accessed online during self-directed learning was significantly correlated with student achievement (r = .79, p < .05). Contribution during reporting phase was measured using the total frequency of relevant concepts uttered during this time. Student achievement was measured by a concept recognition test at the end of the day.

The data were analyzed using a structural equations modeling (SEM) approach. SEM is a statistical technique used to test causal hypotheses among multivariate data. This procedure generates several statistics that enable the investigator to assess how well the empirical data fit the theoretical model and to estimate the strengths of the causal relations hypothesized. In evaluating the goodness-of-fit of the models to the sample data, we used four indicators suggested in the literature: the Cmin/df index of fit, Chi-square, the

Comparative Fit Index (CFI), and the Root Mean Square Error of Approximation (RMSEA) (Browne and Cudeck 1993, Hu and Bentler 1999, Arbuckle 2006). The level of significance (p) computed from Chi-square and degrees of freedom should be higher than 0.05. The Cmin/df index of fit, yielded by dividing the minimum discrepancy (C) by its degrees of freedom should be lower than 3 and preferably close to 1 (Arbuckle 2006). CFI values larger than 0.95 and RMSEA scores below 0.06 can be considered as indicators of good fit (Browne and Cudeck 1993).

Results

Table 3 shows the intercorrelations, means, and standard errors of the variables. There were 35 students from eight teams involved in this study.

(Insert Table 3 here)

The model displayed in Figure 7 was tested against the data, yielding the following results: Chi-square = 12.41, df = 4, $p < 0.05$; the minimum discrepancy, C, divided by the degrees of freedom, Cmin/df = 3.10; the square root of the population discrepancy corrected by the complexity of the model RMSEA = .25; and the Comparative Fit Index (CFI) = .86. These statistics indicate that the model does not adequately represent the data. Figure 8 shows the relevant path coefficients that are statistically significant.

(Insert Figure 8 here)

An inspection of the modification indices and the expected parameter statistics revealed that a slightly modified version of the original model would fit the data much better.

Chi-square for this model is = 3.87 (df = 5, p = .57); Cmin/df = .77; RMSEA = .00; and the CFI = 1.00. Figure 9 shows the relevant path coefficients. Only statistically significant path coefficients are displayed.

(Insert Figure 9 here)

Discussion

The purpose of Study 2 was to further investigate the relationships between the contents of the learning activities of students in a PBL process and their academic achievement. We also sought to investigate the influence of the different learning phases in PBL on student achievement, and how the learning activities of each phase influenced the next. We found a model, based on our findings in Study 1 and on existing theory with regards to PBL (Dolmans and Schmidt 2006, Van den Hurk, Dolmans, Wolfhagen and Van der Vleuten 2001) to fit the data quite well. We will discuss the interrelations found one at a time and relate them to existing theory and empirical findings by others. Subsequently, we will assess the significance of our findings as a whole.

The first relationship is the strong influence of the contents of students' verbal contributions during problem analysis phase on the verbal contributions during the self-directed learning phase. Verbal contributions during problem analysis are likely to reflect students' prior knowledge. However, the extent of students' verbal contribution during problem analysis would also be influenced by students' efforts to activate what they already know about the problem-at-hand (Schmidt 1983, 1992). We know from other studies that prior knowledge, once activated, influences subsequent learning (see for a review: Dochy, Segers and Buehl 1999). This was what was found in experimental studies of PBL as well (De Grave, Schmidt, & Boshuizen, Schmidt, De Volder, De Grave, Moust and Patel 1989, 2001). These studies showed that groups that activated prior knowledge through problem analysis learned more from a subsequently presented problem-relevant text than those who did not. Our original hypothesis was that the extent to which students contribute in discussion during the problem analysis phase would mainly have an impact on the next learning phase- that of self-directed study. However our findings showed that students' verbal contribution during problem analysis influenced not only the SDL phase but also the reporting phase. Hence active engagement in knowledge building and sharing during the problem analysis phase also influences that during the reporting phase. Students who have raised more ideas during the initial knowledge sharing process would perhaps also be more motivated to share and clarify more during the reporting phase, when their knowledge networks are more established. Thus the first phase of the PBL cycle is clearly a very important one since the extent of a student's learning in the PBL process is largely determined at this point. Other factors that could influence this phase include student interest or motivation induced by the problem or the tutorial group itself (e.g. whether it is a productive or dysfunctional group).

Contrary to our original hypothesis though, is that the extent of verbal contributions during problem analysis does not directly influence the amount of individual study in the subsequent learning phase. We will discuss possible reasons for this observation later on.

The next relationship of interest is a relatively strong influence of students' verbal contribution during self-directed learning on the extent of their verbal contribution during the reporting phase. One unique feature of the self-directed learning that occurs in this education context is the collaboration students undertake during this time. While studying individually, students consult each other and interact about the topics at hand. Studies on the self-directed study phase generally investigate only issues related to the individual learning such as time spent on individual study, extent of literature search and extent of studying of literature (Van den Hurk, Wolfhagen, Dolmans and Van der Vleuten 1999, Van den Hurk, Dolmans, Wolfhagen and Van der Vleuten 2001). However in this context, we note that students also discuss and share ideas during self-directed learning, and this collaboration also influences their achievement indirectly through the verbal contributions in the reporting phase. It is quite surprising to note however, that unlike our original hypothesis, there is no direct relationship between verbal contributions during problem analysis phase and the amount of individual study, nor any direct relationship between individual study and verbal contributions during the self-directed learning phase. One possible reason could be due to students who are reserved or are not comfortable to voice out their ideas readily, but do put in significant amounts of individual study during the learning process. Since every student is generally required to share some of their findings and response to the problem during the reporting phase, those who have done more individual study would also be able to contribute more during this phase. Another possible reason for the results that deviate from our original hypothesis could be due to students' perceptions of what makes learning effective in PBL. Although they are not new to PBL, having been in a PBL environment for six weeks of a semester, all of them have come from traditional "teacher-centered" learning environments. Hence there could be significant numbers of students who do not see the benefits of or are not used to the idea of talking and sharing what they have learned spontaneously during self-directed learning time. Another possibility may be that students simply do not wish to share with their peers what they have learned during self-study time. Several studies have shown that peer discussion is important for sharing knowledge and increasing the "collective knowledge of a group" (Rivard and Straw 2000, Hmelo-Silver and Barrows 2008). Some students may not like to share their findings with their teammates, preferring only to share what they have read up on during the reporting phase when the facilitator is present, in an attempt to show that they have done more than their peers. As these are only speculations, further studies are needed to investigate the possible effects of students' perceptions of effective learning in PBL on their learning activities.

Another relatively strong relationship is the impact of the extent of individual study on students' verbal contribution during the reporting phase. A study by Van den Hurk et al. (2001) showed that the quality of student individual study (based on whether learning was done in an 'explanation-oriented way') influenced the depth of students' reporting. This seems similar to our findings where individual study influenced verbal contributions during the reporting phase. However a surprising finding in our context is that there was no direct path from individual study to achievement. Individual study influences achievement indirectly, through verbal reporting. Similarly there is no direct relationship between students' prior knowledge as indicated from their verbal contribution during problem analysis phase to achievement. These findings underline the importance of actively constructing through verbalization for learning to emerge.

Our model thus clearly shows the importance of verbalizations throughout the PBL process. In trying to make sense out of the problem, students produce explanations, initially based on prior knowledge, but in later phases also based on what was learned from fellow students and from the materials studied on the internet. In an earlier study using a similar "online" methodology, we have demonstrated that more than 50% of the learning-oriented verbal exchanges in the small group were collaborative in nature (i.e. involving co-construction, sharing of information etc) (Yew and Schmidt 2008). Visschers-pleijers et al. (2006) similarly found that more than 60% of the verbal interaction during the reporting phase of a PBL tutorial consisted of 'cumulative reasoning' processes. Although no relations to achievement were reported in the studies cited above, Rivard & Straw (2000) showed that giving opportunities for collaborative exploratory talk significantly improved students' post-test scores as compared to the groups which did not have peer interaction. Chi, De Leeuw, Chiu and Lavancher (1994) have also demonstrated that eliciting self-explanations by students results in increased learning. They found that the more students self-explained, the deeper the understanding of the topic that was achieved. Thus the significant role played by verbalizations of ideas in explaining student achievement in our study may be attributed to the PBL approach, which provides opportunities for learning-oriented discussion, and encourages students to verbally interact.

One may argue that students have different levels of knowledge to start with, that those with more prior knowledge tend to contribute more verbally, and that these students would eventually also do better on the achievement test. In this line of thought, the path coefficients found would only reflect that initial differences in aptitude tend to replicate themselves in the different phases of the process and are in itself not an indication of learning. However if this were so, there would likely to be a direct influence of the verbal contributions during the problem analysis phase on students' achievement in the path model. Our results show that this is not the case. Moreover, each phase in the learning process appears to have a unique contribution to learning in the next phase.

Our findings also suggest that learning in each phase of the PBL cycle is a precondition for subsequent learning, thus providing support for the PBL cycle of initial problem analysis, followed by self-directed learning, and a subsequent reporting phase as described by various authors (Barrows 1988, Hmelo-Silver 2004). Moreover, it appears that collaborative learning is to some extent dominant over individual study in predicting students' performance in this PBL context. This can be seen by the strong path coefficients between the verbal contributions at the different phases and eventually to achievement.

The present study has several limitations. First, our structural equation modeling approach has been what is described by Jöreskog & Sörbom (1996) as a 'model generation' approach, where an initial model is fit to data and then modified as necessary until it fits adequately well. MacCallum and Austin (2000) warned in their review that such modifications may sometimes lack validity and are susceptible to chance effects. We recognize that our resulting model is data-driven to some extent, but would also argue that our modifications are meaningful in the context and do not contradict existing theory. The main difference in our modified model is that the role of verbal contribution during problem analysis phase influences not only the next SDL phase but also the reporting phase. Such a model is still in line with our original theoretical basis and also with the findings demonstrating the importance of prior knowledge (reviewed by Dochy, Segers and Buehl, 1999), and that of active integration of new knowledge with prior knowledge (Chi, Deleeuw, Chiu and Lavancher 1994). Second, and related to the first issue, the number of participants for this study (N = 35) was by necessity rather small (although sufficient to demonstrate the expected effects). The data-intensive approach chosen, with its emphasis on the detailed analysis of verbal protocols and internet log files, precluded the involvement of more students. Therefore, we were not able to test our model against a new group to check whether it would survive cross validation. Cross validation is also necessary to check whether our findings are limited to the particular brand of PBL studied or has a broader significance (MacCallum and Austin 2000). The reader is reminded here that PBL as practiced in this study is a specific 'one-day-one-problem' approach. While we have argued that this approach is indeed PBL as it contains all the defining characteristics of PBL (Barrows 1988, Schmidt 1993), the short learning cycle, in particular with the SDL phases being only about 4 hours in total, is likely to have an impact on students' learning processes. Thus our findings in this study may not necessarily be easily extrapolated to PBL in other contexts.

While this study does have several shortcomings, it is a first attempt at developing and testing a model of the process of PBL in an actual educational context using naturalistic data. In particular, we think this study represents the first attempt to provide an insight into the self-directed learning phase of student learning using an analysis of usage of online resources. While we recognize possible limitations to this methodology, our results show that students' learning can be predicted by counting the number of relevant concepts they have accessed during the individual study period. Of course, the use of concept counting only provides an estimate of the quantity of students' learning without providing information regarding the quality or depth of their understanding. We have however shown that this method, while not without flaw, does fairly accurately predict students' achievement. However studies in which students' verbal contributions are analyzed more qualitatively should be carried out to further verify our findings.

General Discussion and Conclusion

Our results from Study 1 indicate that the process of PBL is characterized by two distinct phases: a phase in which there is a high degree of concept articulation and a later concept repetition phase. Thus when viewed as a whole, it can be seen in the initial phase of PBL, students are confronted with relatively large amounts of information from their readings and discussions (knowledge acquisition phase) while the repetition and elaboration of the concepts in the second phase indicate the process by which students organize and perhaps integrate the concepts with their prior knowledge (Mayer 1992). The PBL learning process is therefore one that encourages active processing and organization of information through co-constructions and elaborations in small group discussions.

Secondly, we have also shown that the contents of learning activities undertaken in PBL play important roles in predicting student achievement. Both collaborative learning (verbal interactions) and self-directed study appear to have similar degrees of influence on and importance for students' learning. Moreover, our findings suggest that group processes in PBL provide opportunities for active construction of what is learned during self-study. In order for what is learned during self-study to lead to eventual achievement, verbalization of ideas within the group appears to be essential. We have also demonstrated that the learning in each phase of the PBL process is dependent on the earlier phase, thus providing support for the idea that PBL is indeed a process of sequential steps each building upon the other (Schmidt 1993).

Finally, we have also demonstrated that the recording and analyzing of students' learning-oriented verbalizations and their computer resources log is a useful methodology that provides an account for ongoing learning in the PBL process. Our study has thus provided insights into the active learning within the PBL process through a naturalistic approach.

References

Arbuckle, J. L. (2006). *Amos 7.0 User's Guide*. Chicago: SPSS.

Barrows, H. S. (1986). A taxonomy of problem-based learning methods. *Medical Education,* 20, 481-486.

Barrows, H. S. (1988). *The Tutorial Process*. Springfield Illinois: Southern Illinois University School of Medicine.

Barrows, H. S. (1996). Problem-based learning in medicine and beyond: a brief overview. In L. Wilkerson & W. H. Gijselaers (Eds.), *New directions for teaching and learning* (Vol. 68, pp. 3-11). San Francisco: Jossey-Bass Publishers.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). Newbury Park, CA: Sage.

Capon, N., & Kuhn, D. (2004). What's so good about problem-based learning? *Cognition and Instruction,* 22(1), 61-79.

Chi, M. T. H., Deleeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting Self-Explanations Improves Understanding. *Cognitive Science,* 18(3), 439-477.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher,* 23, 13-20.

De Grave, W. S., Boshuizen, H. P. A., & Schmidt, H. G. (1996). Problem based learning: Cognitive and metacognitive processes during problem analysis. *Instructional Science,* 24(5), 321-341.

De Grave, W. S., Schmidt, H. G., & Boshuizen, H. P. A. (2001). Effects of problem-based discussion on studying a subsequent text: A randomized trial among first year medical students. *Instructional Science,* 29(1), 33-44.

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research.,* 69(2), 42.

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: a meta-analysis. *Learning and Instruction,* 13(5), 533-568.

Dolmans, D., & Schmidt, H. G. (2006). What do we know about cognitive and motivational effects of small group tutorials in problem-based learning? *Advances in Health Sciences Education* 11(4), 321-336.

Dolmans, D., Schmidt, H. G., & Gijselaers, W. H. (1995). The relationship between student-generated learning issues and self-study in problem-based learning. *Instructional Science,* 22(4), 251-267.

Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher*, 23(5-12).

Gallagher, S. A., Stepien, W. J., & Rosenthal, H. (1992). The effects of problem-based learning on problem solving. *Gifted Child Quarterly,* 36(195-200).

Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology,* 40, 631-666.

Glenn, P. J., Koschmann, T., & Conlee, M. (1999). Theory presentation and assessment in a problem-based learning group. *Discourse Processes*, 27(2), 119-133.

Hak, T., & Maguire, P. (2000). Group process: The black box of studies on problem-based learning. *Academic Medicine,* 75(7), 769-772.

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review,* 16(3), 235-266.

Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating Collaborative Knowledge Building. *Cognition and Instruction* 26, 48-94.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling,* 6(1), 1-55.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8 User's Reference Guide*. Chicago: Scientific Software International.

Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., et al. (2003). Problem-Based Learning Meets Case-Based Reasoning in the Middle-School Science Classroom: Putting Learning by DesignTM Into Practice. *Journal of the Learning Sciences,* 12(4), 495-547.

Koschmann, T., Glenn, P., & Conlee, M. (1997). Analyzing the Emergence of a Learning Issue in a Problem-Based Learning Meeting, *Medical Education Online* (Vol. 2, pp. 1-9).

MacCallum, R. C., & Austin, J. T. (2000). Applications of Structural Equation Modeling in Psychological Research. *Annual Review of Psychology,* 51, 201-226.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review* 92(3), 289-316.

Norman, G. R., & Schmidt, H. G. (1992). The Psychological Basis of Problem-Based Learning - a Review of the Evidence. *Academic Medicine,* 67(9), 557-565.

Novak, J. D. (1998). *Learning, Creating, and Using Knowledge: Concept maps as facilitative tools for schools and corporations.* Mahwah, N.J.: Lawrence Erlbaum Associates.

Palincsar, A. S. (1998). Social Constructivist Perspectives on Teaching and Learning. *Annual Review of Psychology,* 49, 345-375.

Rivard, L. P., & Straw, S. B. (2000). The Effect of Talk and Writing on Learning Science: An Exploratory Study. *Science Education* 84(5), 566-593.

Schmidt, H. G. (1993). Foundations of problem-based learning - Some explanatory notes. *Medical Education,* 27(5), 422-432.

Schmidt, H. G., De Volder, M. L., De Grave, W. S., Moust, J. H. C., & Patel, V. L. (1989). Explanatory models in the processing of science text: The role of prior knowledge activation through small-group discussion. *Journal of Educational Psychology,* 81(4), 610-619.

Schmidt, H. G., Loyens, S. M. M., Van Gog, T., & Paas, F. (2007). Problem-based learning is compatible with human cognitive architecture: Commentary on Kirschner, Sweller, and Clark (2006). *Educational Psychologist,* 42(2), 91-97.

Schmidt, H. G., & Moust, J. H. C. (2000). Factors affecting small-group tutorial learning: a review of research. In D. H. Evensen & C. E. Hmelo-Silver (Eds.), *Problem-based learning: A research perspective on learning interactions* (pp. 19-52). Mahwah, NJ: Lawrence Erlbaum.

Schmidt, H. G., Van der Molen, H. T., Te Winkel, W. W. R., & Wijnen, W. H. F. W. (2009). Constructivist, Problem-based, Learning Does Work: A Meta-analysis of Curricular Comparisons Involving a Single Medical School. *Educational Psychologist* (Accepted for publication).

Van den Hurk, M. M., Dolmans, D., Wolfhagen, I., & Van der Vleuten, C. P. M. (2001). Testing a causal model for learning in a problem-based curriculum. *Advances in Health Sciences Education,* 6(2), 141-149.

Van den Hurk, M. M., Wolfhagen, I., Dolmans, D., & Van der Vleuten, C. P. M. (1999). The impact of student-generated learning issues on individual study time and academic achievement. *Medical Education*, 33(11), 808-814.

Visschers-Pleijers, A. J., Dolmans, D., de Leng, B. A., Wolfhagen, I. H., & Van der Vleuten, C. P. (2006). Analysis of verbal interactions in tutorial groups: A process study. *Medical Education*, 40(2), 129-137.

Yew, E. H. J., & Schmidt, H. G. (2008). Evidence for constructive, self-regulatory, and collaborative processes in problem-based learning. *Advances in Health Sciences Education*, 14(2), 251-273.

Zimmerman, B. J. (1990). Self-Regulated Learning and Academic-Achievement - an Overview. *Educational Psychologist*, 25(1), 3-17.

Appendix 1

Basic Science problem that students worked on for the day

Code of Life

I am the family face;

Flesh perishes, I live on,

Projecting trait and trace

Through time to times anon,

And leaping from place to place

Over oblivion.

From "***Heredity***" by Thomas Hardy

(First published in *Moments of Vision and Miscellaneous Verses*, Macmillan, 1917)

The idea of the gene came first. The gene is the thing that carries information about the living organism. The gene tells if one's hair is black and eyes are blue. The gene tells if one can curl one's tongue. The gene carries the 'family face' that goes 'through time to times anon' from mother to daughter, father to son, or the other ways across, over time.

Is the gene a substance you can find in your body, or a kind of a soul-like invisible thing?

Explore the concept of a gene and the role it plays in an organism. Is it possible that the gene is represented by an identifiable molecule, one that is able to carry information akin to a line of code, giving it the ability to execute highly detailed tasks? Determine the qualities such a molecule should have.

Table 1.

Correlation analysis between concept recognition test scores and concepts articulated during the different learning phases (Note: Phase 1: Problem analysis phase; Phase 2: SDL-1; Phase 3: Group meeting with facilitator; Phase 4: SDL-2; Phase 5: Reporting phase)

Total number of relevant concepts articulated								
Phase	1	2	3	4	5	Total for 1,3 and 5	Total for 2 and 4 (SDL)	Total
Concept Recognition Test	-.17	.26	.36	.15	.83**	.51	.20	.41
Number of different relevant concepts articulated								
Phase	1	2	3	4	5	Total for 1,3 and 5	Total for 2 and 4 (SDL)	Total
Concept Recognition Test	.16	.42	.38	.26	.82**	.55	.39	.59*
Number of newly emerged relevant concepts articulated								
Phase	1	2	3	4	5	Total for 1,3 and 5	Total for 2 and 4 (SDL)	Total
Concept Recognition Test	.16	.20	.22	-.21	.25	.44	.06	.56

*p < .05, **p < .01

Table 2.

Correlation analysis between concept recognition test scores and concepts studied from online resources during the different learning phases (Note: Phase 1: Problem analysis phase; Phase 2: SDL-1; Phase 3: Group meeting with facilitator; Phase 4: SDL-2; Phase 5: Reporting phase)

	Total number of relevant concepts read from online resources							
Phase	1	2	3	4	5	Total for 1,3 and 5	Total for 2 and 4 (SDL)	Total
Concept Recognition Test	.62*	.75*	.23	.30	.13	.43	.79**	.68*
	Number of different relevant concepts read from online resources							
Phase	1	2	3	4	5	Total for 1,3 and 5	Total for 2 and 4 (SDL)	Total
Concept Recognition Test	.63	.80**	-.06	.55	.12	.36	.79**	.72*
	Number of newly emerged relevant concepts read from online resources							
Phase	1	2	3	4	5	Total for 1,3 and 5	Total for 2 and 4 (SDL)	Total
Concept Recognition Test	.66	.78*	-.77*	-.11	.06	-.19	.67*	.68*

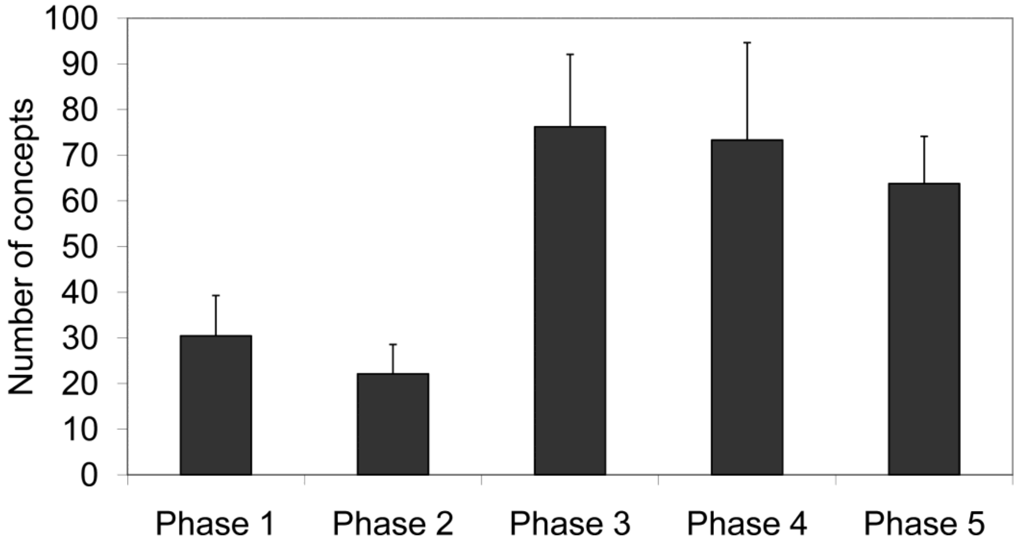
*p < .05, **p < .01

Table 3. Intercorrelations, means and standard errors of the variables (N = 35)

	1	2	3	4	5
1. Verbal contribution during problem analysis	–	.73*	.16	.75**	.29*
2. Verbal contribution during self-directed learning		–	.08	.68**	.23
3. Individual study			–	.32*	.36*
4. Verbal contribution during reporting phase				–	.35*
5. Achievement					–
Mean	16.00	77.23	1039.48	35.17	60.31
Standard error	4.95	15.95	130.33	4.22	2.81

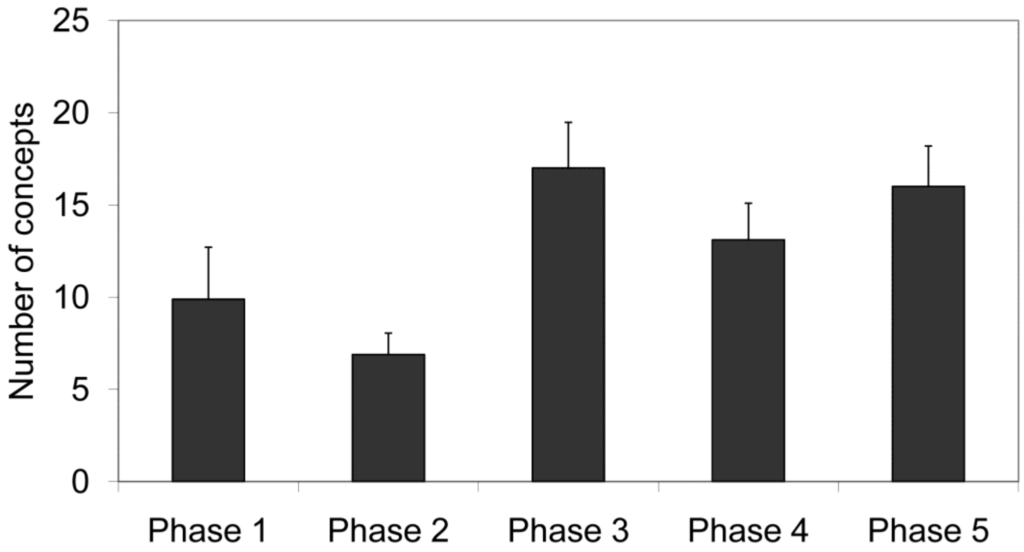
** $p < .01$; *$p < .05$

Figure 1. Distribution of the mean number (+ *SE*) of *total* relevant concepts (includes repetitions) articulated over the different learning phases of the PBL process (N = 9)



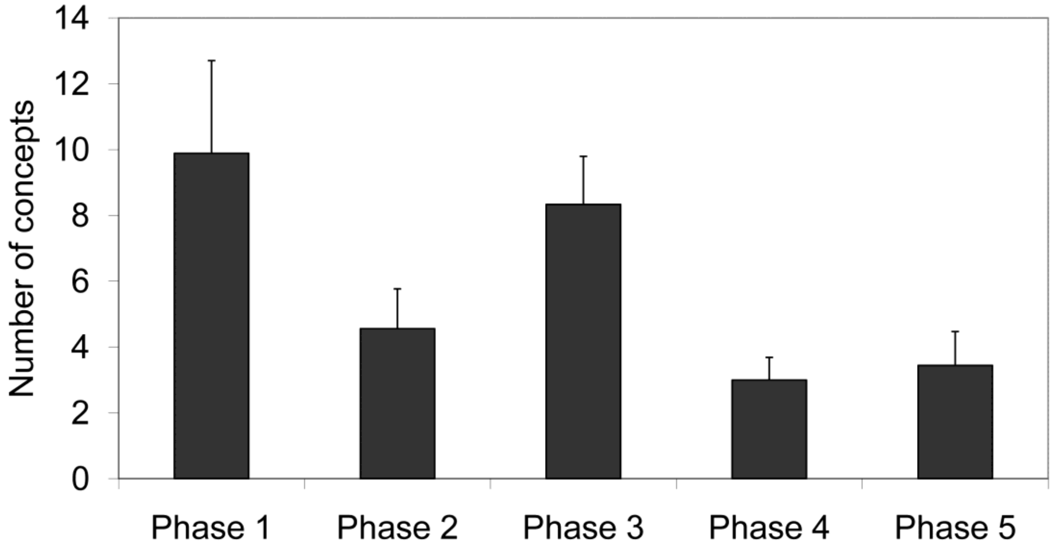
Phase 1: Problem analysis phase; Phase 2: SDL-1; Phase 3: Group meeting with facilitator; Phase 4: SDL-3; Phase 5: Reporting phase

Figure 2. Distribution of the mean number (+ *SE*) of *different* relevant concepts (excludes repetitions within each learning phase) articulated over the different learning phases of the PBL process (*N* = 9)



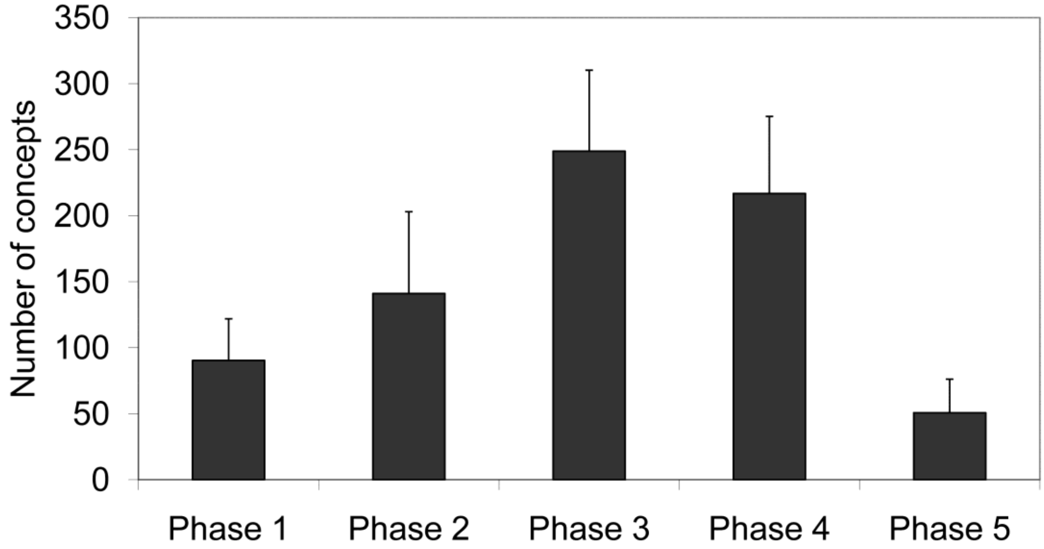
Phase 1: Problem analysis phase; Phase 2: SDL-1; Phase 3: Group meeting with facilitator; Phase 4: SDL-3; Phase 5: Reporting phase

Figure 3. Distribution of the mean number (+ *SE*) of *newly emerged* relevant concepts articulated for the first time in the day over the different learning phases of the PBL process (*N* = 9)



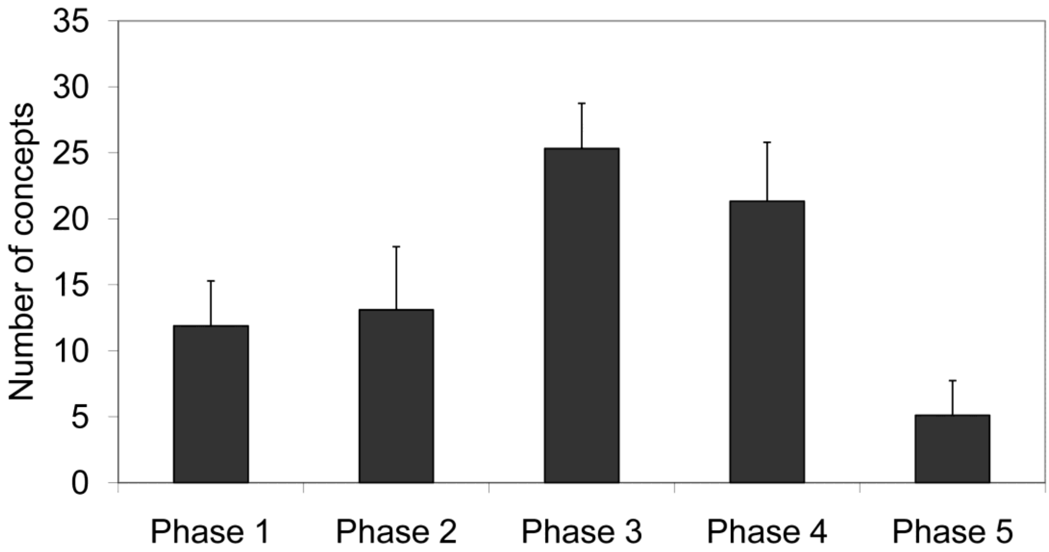
Phase 1: Problem analysis phase; Phase 2: SDL-1; Phase 3: Group meeting with facilitator; Phase 4: SDL-3; Phase 5: Reporting phase

Figure 4. Distribution of the mean number (+ SE) of *total* relevant concepts (includes repetition) accessed online over the different learning phases of the PBL process (N = 9)



Phase 1: Problem analysis phase; Phase 2: SDL-1; Phase 3: Group meeting with facilitator; Phase 4: SDL-3; Phase 5: Reporting phase

Figure 5. Distribution of the mean number (+ SE) of *different* relevant concepts (excludes repetition within each learning phase) accessed online over the different learning phases of the PBL process (N = 9)



Phase 1: Problem analysis phase; Phase 2: SDL-1; Phase 3: Group meeting with facilitator; Phase 4: SDL-3; Phase 5: Reporting phase

Figure 6. Distribution of the mean number (+ SE) of *newly emerged* relevant concepts accessed online for the first time in the day over the different learning phases of the PBL process (N = 9)

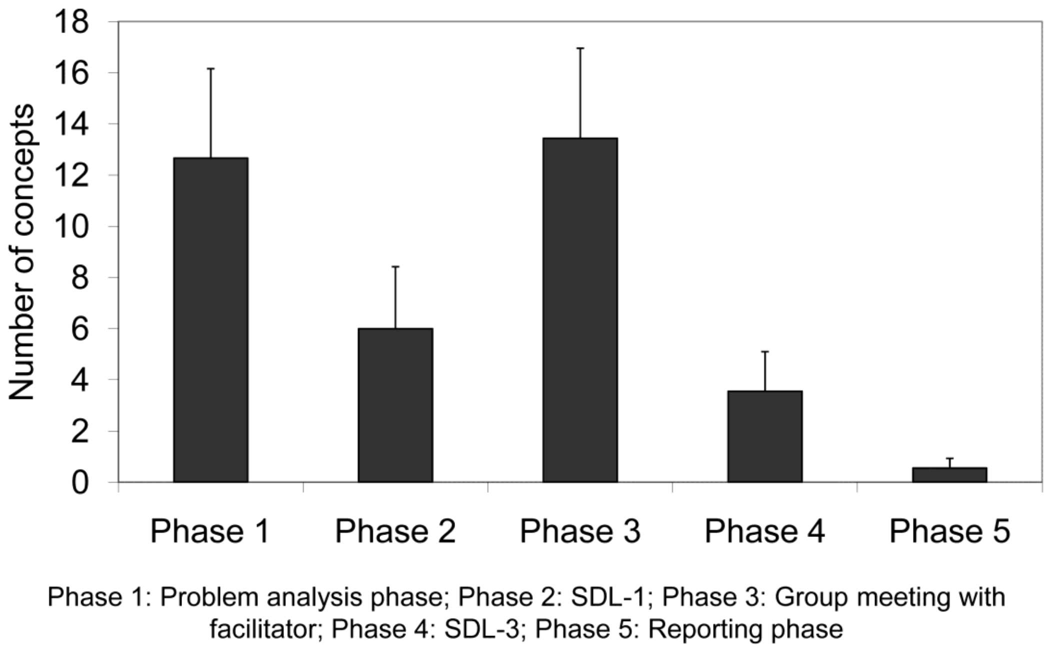


Figure 7. Theoretical model on the learning processes involved in PBL

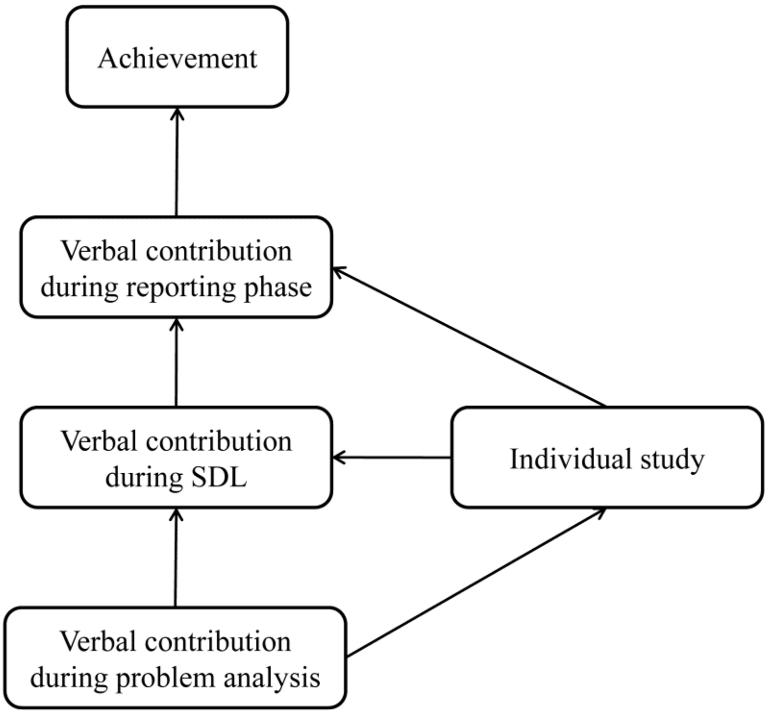


Figure 8. Path model of the hypothesized model on the learning processes involved in PBL (error terms are omitted for readability and only statistically significant path coefficients are displayed)

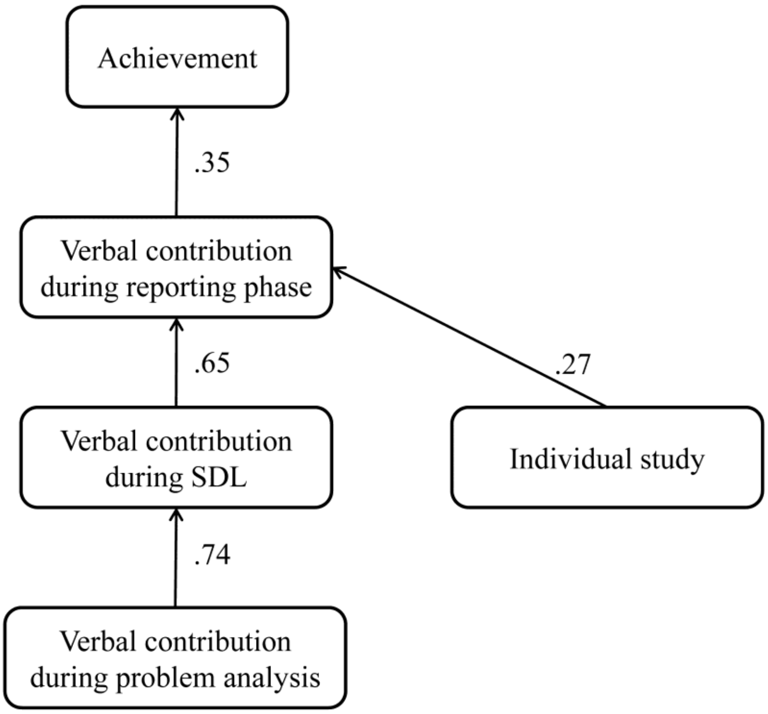


Figure 9. Path model of the learning-oriented activities affecting student achievement